AHRQ's Community Forum

CENTER FOR MEDICAL TECHNOLOGY POLICY

Integrating Stakeholder Preferences in Comparative Effectiveness Research: Potential Uses of Multi-Criteria Decision Analysis Techniques

August 27, 2012

Contents

Webinar

[START RECORDING]

PENNY MOHR: Good morning. This is Penny Mohr. I'm Vice President of Program Development at the Center for Medical Technology Policy. It's my pleasure to welcome you to our webcast sponsored by the Agency for Healthcare Research and Quality effective programming.

The webinar today is entitled Integrating Stakeholder Preferences and Comparative Effectiveness Research Potential Uses of Multi-criteria Decision Analysis Techniques.

We have a wonderful turnout today with over 150 people registered. We are looking forward to a great webcast.

Our agenda for today's webcast will be an interactive presentation on the use of formal structured approaches for eliciting and quantifying stakeholder preferences in comparative effectiveness research. These formal approaches fall under the general categories of multi-criteria decision analysis and stated preference techniques. Two of the specific techniques that we will be comparing and contrasting are analytic hierarchy processes and conjoint analysis. These techniques originated in the fields of behavioral psychology and market research but have been used increasingly in health services research.

Some of you listening today may be familiar with their use for eliciting preferences among treatment options and shared decision making or measuring patient preferences in cost effectiveness analysis. The premise behind today's webcast is that eliciting stakeholder preferences is a fundamental aspect of comparative effectiveness research.

We will be presenting two examples. How they can be used in setting research priorities and how they can be used to help design research protocols.

A brief word about the sponsorship for today's webcast. It was funded as part of ARHQ Community Forum Project for which I am a subcontractor. The Community Forum is charged with working with ARHQ's effective healthcare program components to expand stakeholder involvement in research processes and activities. By stakeholders we mean the

broad array of groups with a vested interest in the outcomes of comparative effectiveness research such as patients, clinicians, payers, policy makers and healthcare administrators.

The Community Forum is also developing methods for gathering public input based on value-based healthcare questions. That said the presentation and panel discussions that we are going to have today do not represent official policy of either ARHQ or the U.S. Department of Health and Human Services.

I also want to note that today's webcast is being recorded and it will be made available on ARHQ's website for your review or for you to pass on to those who were not able to attend today.

Before we begin I would like to set the stage for why we felt this was an important webcast to host. There has been an increased emphasis on comparative effectiveness research in recent years. While there are several definitions of this word many people focus on the comparative aspect. That is comparing evidence on the effectiveness, benefits and harms of different treatment options. I think what distinguishes comparative effectiveness research from the health outcomes or health services research that we have been doing for many years is the focus on its purpose as defined by the Institute of Medicine, shown here.

The purpose of comparative effectiveness research is to assist consumers, clinicians and policy makers to make informed decisions. The hypothesis is that this new stakeholder or patient-centered approach to research will increase the relevance of evidence so that it is more informative to patients and their caregivers and ultimately will improve the appropriateness of healthcare use and patient outcomes.

We approach the webcast today with the belief that these formal methods can be used to elicit stakeholder preferences across the many ways of conducting comparative effectiveness research, including informing researchers about specific research gaps, to guide systematic reviews or aiding investigators in selecting specific features of a study protocol for randomized controlled trials or observational research.

They also can be used throughout the spectrum of activities related to the conduct of comparative effectiveness research, from studying research priorities to assisting in the translation and dissemination of the study results at the completion of research.

For today's webcast we will discuss and include the use of analytic hierarch processes to elicit stakeholder preferences to set research priorities and also the use of conjoint analysis to elicit patient preferences to design research protocols.

I am thrilled today to introduce or panelists. We have an absolutely expert and excellent panel joining us today. Dr. Maarten IJzerman is Chair of the Department of Health Technology and Services Research at the University of Twente in the Netherlands. He is published widely on the topic we are presenting today. He will provide an introductory overview of the use of multi-criteria decision analysis and stated preference techniques for health services research, comparing and contrasting the analytic hierarchy process with conjoint analysis.

Dr. Jerry Krishnan is Professor of Medicine and Associate Vice President for Population Health Sciences in the Office of Health Affairs at the University of Illinois. He is also the Chair of the Steering Committee for the COPD Outcomes-based Network for Clinical Effectiveness and Research Translation, which is known as CONCERT. That is funded by ARHQ to develop a stakeholder-supported research agenda in support of COPD. He will be talking about his experience in using AHP, analytic hierarchy process, to identify research priorities in this area.

Dr. Reed Johnson is a Distinguished Fellow and Principal Economist at the Research Triangle Institute. He helped pioneer the development of basic non-market valuation techniques such as the ones we will be discussing today in the fields of health and environmental economics. He also chairs the International Society for Pharmacoeconomics and Outcomes Research Conjoint Analysis Experimental Design Task Force. He will describe the application of conjoint analysis for eliciting patient preferences for outcomes to design a study protocol.

Here is our agenda for today's webcast. I'd like to point out that there will be a brief question and answer period at the end of each presentation and a broader question and answer period at the end of the session. We will have about five minutes at the end of each of the presenters and then potentially about ten minutes at the end. We hope to have enough to take everybody's questions. I'd like to remind everybody in the audience that you can email us your questions any time by entering your question in the chat box, which is on the left-hand side of your screen and submitting it. We are going to do this in an interactive fashion and we are going to have a few questions scattered throughout the webcast to assess your familiarity with and interest in using these techniques for your own research.

Just to start off the meeting today we have these questions here. How familiar are you with preference elicitation techniques such as analytic hierarchy process and conjoint analysis? If you could just take a minute, fill in your responses there and then submit, we should be able to see the results in just a minute. All right, well, it looks like we have an audience that hasn't had a lot of experience with these techniques. About 40% have never heard of MCDA before so we look forward to giving a good introduction to these techniques and with that I would like to thank everyone for joining us and introduce Dr. Maarten IJzerman to provide you with an introduction to multi-criteria decision analysis techniques and conjoint analysis. Maarten?

Integrating Stakeholder Preferences in Comparative Effectiveness Research Using Multi-criteria Decision Analysis (MCDA) and Conjoint Analysis (CA)

MAARTEN J. IJZERMAN, PhD: Okay, well, thank you, Penny, for your nice introduction and for putting together this webinar, which is really a privilege to be on this webinar. I would like to talk about the integration of stakeholder preferences in comparative effectiveness research using multi-criteria decision methods and conjoint analysis methods, although, technically formerly conjoint analysis is not an MCDA method we are very happy to introduce both of the methods for the audience today.

Can I get the next slide, please? I want to follow up on what you just introduced and in comparative effectiveness research we wish to make an informed decision based on the available clinical evidence for multiple endpoints as collected in evidence reviews and systematic reviews. However, the primary endpoints and the other endpoints chosen in clinical trials may not be the most relevant endpoints for patients and other stakeholders and also process related factors do contribute to the actual use of healthcare technologies by stakeholders and are usually neglected in decision making. There is good reason to incorporate, integrate stakeholder preferences in the decision making process. Actually, there are three different approaches that we can take to include the stakeholder preferences. First we can advise stakeholders to be represented in stakeholder appraisal committees like being done in NICE appraisal committees or with ARHQ on the FDA but in here today we would like to talk about a formal elicitation of preferences to guide the deliberative process so we can come up with a rank order of the multiple endpoints according to the stakeholders and we can use that in our deliberative process decision making and the second approach that we can take is to formally integrate preferences, quantitative preferences, in a decision framework. The two methods that I am going to introduce actually I use to formally elicit preferences to guide the decision making approaches.

Can I get the next slide, please? Both methods, both MCDA and conjoint analysis methods, decompose a decision problem into a set of criteria, sometimes called attributes or levels. Suppose this example if you want to get the best treatment we have to choose for best treatments we can do decompose our decision problem into three attributes. It is the clinical outcome, the benefit of a drug, if there are adverse events or copayments of a particular drug. The three main criteria are the three, clinical outcome, adverse events and out of pocket costs. Then we have to find three different drugs, drug A, B and C, and they all perform differently on each of the endpoints. In terms of clinical outcome drug A is a complete relief of symptoms, performs better than the other two drugs, B and C, so they perform differently. The question here is how do we make a decision and how can we support that decision making process? The first that we do after decomposition of the

problem is that we want to get a weight for each of the criteria.

We go to the next slide you can see I gave some hypothetical weights for each of the attributes involved in our decision making. I gave 0.5 for clinical outcome, 0.33 for the adverse events and 0.17 for out of pocket costs in terms of the overall weight of our decision making framework. The second thing that both methods do, MCDA methods and conjoint analysis methods, is that they use a linear additive value function to determine the relative preference for the alternatives. In terms of the performance we go to L-11 criteria in drug A performs best compared to Drug B and C so it gets a weight of 0.5 compared to 0.25 for drug B and C and here we can use the linear additive value function to come up with a relative preference for each of the drugs and I gave the example for drug A. It will give you the overall relative preference of 0.375.

We go to the next slide. Multi-criteria decision methods and conjoint analysis they differ in the approach that they take in studies to weigh each criteria. MCDA methods, multi-criteria decision analysis methods, is a sub-discipline of operations research for decision sciences and explicitly considers multiple criteria in a decision making framework. MCDA methods enable the evaluation of many alternatives by explicit ranking or rating or pairwise comparison of the criteria and the alternative. I am going to talk about the Analytic Hierarchy Process, which is one of the most widely used MCDA techniques available.

The other approach, conjoint analysis, offers subjects a series of choices among two or more product profiles. It actually generates a hypothetical market in which consumers may choose the best options. The pattern of choices reveals the implicit decision weights of patients attached to a therapeutic benefits, harms, processes and costs that describe the treatment profiles. That is the definition that comes from the ISPOR taskforce, which is chaired by Reed Johnson.

We go to the next slide. I will give you an example so suppose that we do a study in which we have to select a good restaurant for dinner for tonight and we have defined three main criteria, a cooking style, travel distance and

price, and we have three different levels for each of the criteria so for cooking style we can choose between Italian food, Greek or Thai food and likewise we can choose for differences in travel distance and the price attributes.

Conjoint analysis will generate scenarios based on the levels. In this case you can generate 27 different scenarios that describe all combinations in the decision space, the hypothetical market. We can generate all the options with 27 different scenarios.

MCDA will ask you which criteria and which levels are important using ranking, explicit ranking of different criteria, rating or pairwise comparisons. In the case of pairwise comparisons 12 different comparisons can be sufficient to describe the complete decision space so we first can compare the three main attributes, cooking style, travel distance and price, and then we can compare the different levels, Italian food, Greek food or Thai food. Basically, that is the difference between both methods.

If you go to the next slide you can see what this will look like. This is an example of a choice set, which is used in conjoint analysis, so we offer each participant a number of choice sets that are based on the level distribution as I just showed you so we have three underlying levels in our hypothetical market, the food, the travel distance and the price, so we can generate a number of choice sets and they are offered to respondents, to the stakeholders, and based on their response it is possible to estimate the part-worth utility, the value, the weight of the levels and attribute weights, an indirect estimation of the attributes.

For instance, if you present Thai food for a number of times in this scenario and the respondent takes all the scenarios we can assume Thai food is an important factor in our decision making.

If you go to the next slide you can see a decision structure as we develop for the AHP, which is one form of multi-criteria decision analysis and Jerry Krishnan will talk about AHP as an example. Here we choose the best restaurant. We decompose our decision problem. We have three main attributes. The style of the food, the price and the travel and then with an other series of sub-criteria and here we have the same level weights that we

have to establish by using pair-wise comparison outcome. I will show you in a minute how we can come up with pair-wise comparisons and weights for each of the attributes in an AHP approach. What are the differences between our methods?

We go to the next slide and there are quite some differences and I will not explain all the differences. It is very important to understand that AHP and MCDA techniques to help to improve judgment they start providing decision support to a decision maker, whereas, conjoint analysis methods were developed in marketing and psychology and behavioral economics to imitate consumer judgment so we generate a hypothetical market with scenarios, with options that people have to choose from.

The conjoint analysis methods, because there are multiple attributes that you have to simultaneously, are considered to be more cognitively stressful compared to AHP. On the other hand, because we have decomposed decision approach where we have pair-wise comparisons of attributes one at a time AHP and MCDA may be less realistic compared to conjoint because in the real world life we make a decision based on a whole scenario, a choice set, really.

Finally, very important to know is that in doing a conjoint analysis method because you have an indirect estimation of your attribute weights you need large data sets with lots of data from respondents, whereas, in MCDA approach, like AHP, you can even do that in a single person so you can provide weights from a single individual doing all the pair-wise comparisons.

Okay, we go to the next slide. This is the basic structure of an AHP approach so we have a decision objective, the selection of a treatment with a series of criteria that we make our decision on, base our decision on. We can use sub-criteria and then we have a couple of alternatives to our treatment plans. Each of these levels sum up to one so criteria one, two, three the weights sum up to one. How do we get the weights in AHP?

We go to the next slide. We do that with pair-wise comparison and this is an example of a pair-wise comparison in a matrix with four criteria. This is a hypothetical example where we have four criteria, clinical benefit,

impact of the treatment, side effects and additional cost to patients. We ask people to weigh each of the criteria using a pair-wise comparison so you see on the right side a decision matrix, four criteria, there are some wide spots there in the matrix, four criteria can be established using six pair-wise comparisons. We have to make six pair-wise comparisons to complete the full matrix. We ask them the question how important is each criteria compared to the other? You can see the verbal scale of importance in the bottom of the slide. One means equal so there is no difference between the two attributes. We can go out to nine, which means that the one attribute is extremely more important than the other attribute.

We go to the next slide I will briefly show you what you can do. We have constructed the matrix of pair-wise comparisons, which is here in the green side of the slide; it is right upper part of the table. This is the original data included from a respondent so then we take the reciprocal scores on the left lower part of the table and we can complete our entire matrix. AHP uses a method called Eigenvalue to come up with a priority score so the first step to take is to sum each of the criteria so we take a sum score for each column here, which is presented in the lower end of the slide, and we can use a sum score to generate normalized scores.

If you go to the next slide you can see the normalized scores and the original data includes divided by the sum score it will give you the normalized score and then we take th average for each row, which will give you the overall priority score for each of the criteria so in this example the clinical benefit will get a weight of 0.49 and these weights do sum up to 1.0. This is basically how we use the Eigenvalue to generate the weights in an AHP approach.

If we go to the next slide then you can see we can use, as I said, we can use it in a single person, a single patient or stakeholder, but one very big advantage of this kind of MCDA approach is that you can also use it in a group decision support system so we can invite a panel of experts, a panel of patients or stakeholders to discuss the decision problem at hand and to provide their scores on the screen so everybody gets a remote unit. They provide their scores and the facilitator can share all the scores on the

screen and if there is a lot of dis-concordance in the panel we can allow discussion and see if people can build a consensus on their decision problem. Here we used the same approach with a pair-wise comparison in a matrix of four criteria and we asked our panel to come up with this course and build a consensus. It is really useful for consensus building and guideline development. We have used this kind of accrued physician support in a study in Germany.

If we go to the next slide where we worked with the Institute for Quality in Healthcare in Germany in a case of antidepressants. We used formal benefits assessments reports from e-quick [phonetic] in Germany to establish all of the endpoints that were used in clinical trials and systematic reviews and we tried to build a decision structure based on all of these endpoints. We came up finally with a decision structure with three main criteria, the efficacy, adverse events and disease-specific quality of life and then we defined another series of sub-criteria. If you look at efficacy of the antidepressant treatment, we looked at the response of the drug, the remission and whether there would be a relapse of a drug. Then we invited two panels, one panel with patients and one panel with psychiatrists.

If we go to the next slide you can see the results of our panel sessions because we asked both panels to complete this decision structure in terms of the weighing. Here you see the weights obtained from the panel from patients and from the psychiatrists. If we down the left side you see that the main attributes, effectiveness here, sorry, the previous slide, effectiveness, you see subdivided to response, remission and relapse. You see there is a clear difference to how psychiatrists and how patients respond to the endpoints in terms of importance. Where patients think that even immediate response of a drug because they are in search of a cure physicians, psychiatrists, think that the final remission of complaints is much more important in their decision making so there is a clear dis-concordance between what patients think and what psychiatrists think. You see also the other criteria are less important in their decision making. This is the formal ways of elicitation for the criteria used in our decision structure but then we go to the next slide.

The second step is to integrate the clinical evidence so here you see in this figure the bars in the bottom of the figure represent the weights, the decision weights for each of the criteria. This is the data from the psychiatrists because response is more important than remission in their weight, 0.30, and the colored lines you see the performance scores. We tested for three different antidepressants and we compared the three different antidepressants on each of the criteria. In terms of response to a drug venlafaxine, the blue one, the blue color, outperformed the other two drugs so this falls better in terms of response to drug. That will generate the overall preference function using this added addition. Here you can integrate the clinical evidence and the weights obtained from stakeholders and in this case it was patients and psychiatrists.

Of course, you can use clinical evidence from clinical trial data so we can transform from large clinical trials into a linear score as we can use here in an AHP approach.

We go to my last slide because I show you a couple of advantages of using MCDA and conjoint analysis methods and Jerry Krishnan and Reed Johnson will continue with that after my talk and I think I would like to conclude with a couple of general remarks. I think both methods, MCDA and conjoint analysis methods, stated preference techniques do support decision making but they do not make a decision. Also, conjoint analysis methods, also called stated preference techniques, are used to obtain stakeholder preferences over a wide range of treatment options and MCDA techniques, like AHP, as I showed you, can also support a process of decision making, particularly using this group decision approach. The value of MCDA is not the decision algorithm itself. It doesn't really quantify a decision but the advantages make the process of decision making more explicit and more transparent, particularly to your stakeholders. Preference data, finally, is to be used in conjunction with clinical data. Preference data can never replace original clinical evidence. Thank you very much. Penny, back to you.

MS. MOHR: Maarten, that was fantastic. Thank you. It was really terrific and clear presentation. I would like to open it up for questions now and we actually do have one question. Tanya [phonetic], if you could go back to slide 12 for us there was a specific question about that slide.

I think it is this one right here on the Eigenvalue. The specific question was is there a typo in the five versus one-fifth in the notation? I guess this is speaking about the notation on the adverse events. Is that right? I'm not exactly sure.

DR. IJZERMAN: The question is is there a typo if there should be a five instead of a one to five by five? A five I can see very quickly. I don't see the typo. Anyway, if there is a typo it should be the reciprocal score so if it is a three let's go to the first example. The clinical benefit compared to adverse events, which is agreeing. It is a three. It should be referred to the one divided by the three so you get one-third of the reciprocal score so that should be the reciprocal here. If there are some typos I can't see it here at the moment.

MS. MOHR: Okay. Thank you very much. Does anybody else have questions? You can submit them on the chat box on the left-hand side and just press submit. We will give a few minutes for people to see if they have any other questions. There is a question here from one of the participants whether or not there are specific methods to select sample sizes for MCDA?

DR. IJZERMAN: That's an excellent question. Thank you for repeating the question. As I showed you, you can use MCDA in a single stakeholder so that is what we have used for shared decision making problems, for instance, so we can implement a web-based approach using AHP ways. We don't need any sample sizes here. In general, sample size has to do with the precision of the estimate score, the weights, and I am not sure if there is a rule of thumb. There is for conjoint analysis, of course, and Reed will talk about them, maybe. There is, of course, the one we are concerned with is the precision that we have for our weight estimation and of course the more people we have the more precision that we have. The other thing that is really relevant to consider in the MCDA is the heterogeneity so if we use MCDA approaches in the large survey and we survey among a large group of people there might be preference heterogeneity within a sample so different people with different ages or backgrounds may have different preferences. We have to tackle that difference and I think that is more important to see that preference heterogeneity than the actual statistical precision in their estimates.

MS. MOHR: I guess, Maarten, I have a question, then, because conjoint analysis does require a fairly large number of respondents and how do you calculate sample sizes for conjoint analysis?

DR. IJZERMAN: In conjoint analysis we usually use a rule of thumb that really depends on the number of attributes that you have, the number of levels within each of the attributes, the number of choice sets that you can offer. That determines the sample size. I think Reed will talk about it or maybe can answer to that later on. There is a rule of thumb and if people don't know that rule of thumb we can provide it, of course. I just see on the chat box that there is a typo in this slide so, Jerry, thank you for correcting for me. The one shouldn't be five there. It should be one divided by five so thank you for noted that. There is a typo.

MS. MOHR: Okay. That's great. I think we have one more question. I am not exactly sure if I understand this question but one of the participants asked if a quality of life survey plays one of the main roles in this technique?

DR. IJZERMAN: What we do in MCDA we can incorporate quality of life surveys and we can even incorporate questionnaires and we can transform question to the weight, a priority score, and quality of life in the example I showed you in the German case on antidepressants it is one of the main criteria that we use for decision making. Here we can attach a weight to quality of life.

MS. MOHR: I guess I will ask one more question here and then we are going to have to save the rest of these questions for later. There are some really good ones that are coming up right now but there is a question here about whether there is a difference between resources in team members that are needed for each approach?

DR. IJZERMAN: That is also an excellent question. There are different approaches so we can take a shared decision making approach as a single person that we guide for an individual decision making process and you will just need a computer with the software and it will support your decision making. That is pretty easy. The other approach is the group decision making approach and we usually invite panels over 10, 15 or 20 people at the same time and we

have a lot of preparatory work developing and constructing the decision tree and the decision construction and we discuss that with the panel first. Obtaining the preference weights of our panel usually doesn't take more than a day, between 4 and 16 hours, to get the really weights for the pair-wise comparisons in our AHP structure but there is a lot of work to be done before coming up and inviting the panel. Finally, of course, we can survey, we can obtain preference weight in a survey-based approach and that also requires a lot of survey, of course, and what best survey that we can administer. Even here the most of the work has been done in preparing the survey, developing the decision tree, being very clear about which criteria to include and which not so that is the resources that really takes the most time.

MS. MOHR: Thank you very much, Maarten. I am going to move on now to Jerry Krishnan. Oh, I'm sorry. I apologize. I forgot that we had a polling question here. The polling question is how likely would you be to consider using preference elicitation techniques in your own work? If you can just spend a moment and answer whether or not it would be not at all likely, somewhat likely or very likely we will give you the results very shortly. Well, this is very promising. Looks like well over 80% of the people say that they would be somewhat likely or very likely to use these approaches so fantastic, Maarten, I think you have also done a great job. On to Jerry, now, and thank you. Jerry?

Use of Analytic Hierarchy Process to Elicit Stakeholder Preferences for Prioritizing Research

JERRY A. KRISHNAN, MD, PhD: Good afternoon everyone. Thank you very much to AHRQ and Penny for providing me the opportunity to discuss with you the experience of CONCERT in identifying research priorities for COPD. I will be discussing the use of analytic hierarchy process, the stage of which was set very nicely by Maarten here, to how we use it to elicit stakeholder preferences for prioritizing research. I think what you will see here is that this was a journey for the CONCERT group. We actually used a variety of different methods that ultimately moved towards use of AHP because of some of the natural advantages it offers.

Next slide. Here is the outline for my talk. I will begin with a very brief summary of why chronic obstructive pulmonary disease as a key health condition. Again, as part of the journey I will share with you our experience in using simple rating mechanisms such as asking stakeholders to vote on the importance of specific topics. I will also review with you our experience when asking stakeholders to simply rank order various topics and then I will close with our experience in using analytic hierarchy process and tell you a little bit about how we modified a little bit the processing in order to make it practical in this application when looking at stakeholder priorities.

Next slide. I am going to begin by saying that chronic obstructive pulmonary disease is a key health condition in, of course, not only for the U.S. but also for much of the world. In the U.S. it is actually the most common lung disorder. It is now the third leading cause of death, surpassing cerebrovascular accidents about two years ago. It is also the third leading cause of hospital readmissions and with healthcare reform on its way COPD is trying to attract more attention as an opportunity for comparative effectiveness research and reporting outcome and lastly it is a very expensive condition and in the U.S. alone costing about $50 billion a year.

Next slide. COPD is not only important because of its attributes I just described but also many believe that it represents a model health condition for studying complex medical disorders. Patients with COPD often have other comorbid conditions, be it mental health conditions or cardiovascular disorders. In fact, about 95% of patients with COPD also have another clinically important comorbid condition. Patients with COPD also tend to be receiving care from a variety of healthcare providers, nurse, respiratory therapists, physicians, and even among the physicians tend to cycle between specialists and primary care physicians providing lots of opportunities to understand how healthcare is delivered as you cross multiple healthcare providers. Finally, patients with COPD tend to cycle between acute care and chronic care settings. In the U.S. alone, for example, there is about 500,000 to 750,000 hospitalizations for COPD exacerbations each year and the majority of those patients then go on to home after hospital discharge or long-term care facilities providing

opportunities to understand better how to improve care across transitions in care.

Next slide. The COPD Outcomes-based Network for Clinical Effectiveness and Research Translation received funding from the Agency for Healthcare Research and Quality in the U.S. to identify the effectiveness and translational research priorities to improve COPD care. In year one we focused on a chronic COPD care that is outpatient care or coordinating care across healthcare providers and healthcare settings or otherwise known as care coordination. In year two we then focused on acute COPD care so the care providers in emergency departments and hospitals as well as in transitions in care as patients cycled between acute and chronic care settings. Because the engaged stakeholders were over a two-year period across these various research areas we also had the opportunity to examine different approaches to setting stakeholder priorities for research. In year one we tended to focus on simple important scores and use of rank and, again, because of some lessons we learned in year two we then moved towards the use of analytic hierarchy process.

Next slide. Who is involved, what did the stakeholders do and what sequence did the stakeholders perform these various activities? We engaged a very diverse pool of stakeholder groups, examples of which are provided here on the slide. In total we had about 61 stakeholder organizations involved in the prioritization process over a two-year period.

Next slide. The stakeholders were engaged over, again, a two-year period and involved three distinct phases. Initially we had pre-conference teleconferences, potentially, where we set the goals and procedures for what we would be doing. We used these engagements to elicit topics for research and we conducted provisional voting in order to queue up what we do when we would meet in person. During in-person meeting we had presentations by topical experts or content experts in order for the stakeholders all to get on the same page before they would then vote. We then had discussions about the provisional votes highlighting where different stakeholder groups identified different priorities we used this opportunity during the in-person meeting to have some discussion so we could better understand why different stakeholder groups felt

differently about research priorities. Then, we conducted some final voting procedures, which I will describe momentarily. Post-conference is just really where we are now. We then developed a report, submitted the report to all the stakeholder groups and invited comments and our report now is in peer review and we are hoping that we will get the good news soon here.

Next slide. What I am showing you here are some examples of what we found in year one. We asked stakeholder groups to rate the importance of various topics that were proposed for chronic COPD care. For example, an example of the findings when we asked stakeholders to rate the importance of various topics is shown here.

Next slide. What we found is that a number of stakeholder groups, again, we had 61 different stakeholder groups identify 9 broad topical areas, represented here. We found that stakeholders had very distinctly different preferences for these various topics regarding its importance. This is shown here by the wide interquartile ranges. You can see, for example, that topic A had an interquartile range of importance from one to three; topic H had an interquartile range ranging from five to ten. We also found that simply asking stakeholders to rate the importance of topics was insufficient. It didn't really provide much separation across different topics. As you can see here, several topics all had a very similar importance score, for example, three. We also found that simply asking stakeholders to rate the importance of topics didn't provide information about the rationale or criteria for voting that way, hence some of the limitations when simply asking for importance of topics.

Next slide. Shown here is what we found when we asked stakeholders in a separate meeting to rank order topics. For example, rank ordering from one to nine if you had nine topics here. While we found that there were distinct limitations when asking stakeholders to rate the importance of topics we found also that simply rank ordering topics also had its limitations.

Next slide. Again, in this case I'm giving you an example of nine topics. There were 9, 10 and 11 topics depending on which area we asked stakeholders to nominate topics for. You will see here, again, there was wide confidence

intervals or I should say interquartile ranges indicating variable preferences across different stakeholder groups. One of the key limitations we found when simply rank ordering topics from the most important, let's say, to the least important from one to nine is that simple ranks don't really provide a measure of the relative importance of topics. Take, for example, the case of topic A, which had a rank order of three so the median importance was felt to be three. This next highly topic was topic B that had a rank of 3.5. It is unclear, for example, how much more important topic B is relative to topic A even though topic A has such a higher rank. Said another way, simply rank ordering the topics doesn't quantify the relative importance of topics even though you can put them in some rank order. Then, finally, as in described in the previous slide when looking at the relative importance of topics rank ordering topics this way doesn't provide any linkage to the criteria used by stakeholders in eliciting voting preferences.

Next slide. I am going to present to you now the third and last part of the talk, which has to do with the use of analytic hierarchy process here. I borrowed liberally one of Maarten's earlier slides and one of the advantages we found in the analytic hierarchy process is that it explicitly links treatment alternatives or choice alternatives to the criteria shown here. One of the other points to make is that the formality hierarchy process is a quantitative measure, sometimes called normalized priorities, which represents a proportion of the total importance that is attributed to a particular decision alternative.

Next slide. Let's look at an example here when we apply analytic hierarchy process to different research topics. As described by Maarten earlier analytic hierarchy process explores a series of pair-wise comparisons between different alternatives. In this case now it is no longer endpoints here. I am showing you different topics, topics one through four, and as they relate to each of the criterion listed above.

Next slide. For example, if stakeholders are asked to vote on which the extent of a research topic, one, and in this case research topic two, meet criterion one using a range of values from one-ninth to nine so for example if you

believe that topic one is five times as likely to meet criteria one as topic two you would give topic one versus two a score of five.

Next slide. You will then proceed on to the next pair-wise comparison, which is comparing topic one to topic three and again you will go through a similar situation in which you will ask the stakeholder to rate topic one versus topic three and providing some quantitative measure about the extent to which topic one versus topic three meets criterion one.

Next slide. Similarly, you will repeat the process for topic one versus topic four.

Next slide. Because you are asked to do a series of pair-wise comparisons you will also need to compare topic two to topic three.

Next slide. And topic two to topic four.

Next slide. And topic three to topic four.

Next slide. As you can see here when you have four alternatives, let's say four topics, and you have a single criterion in which you are asking stakeholders to evaluate for relative importance the topics will have to do six pair-wise comparisons before, topics, for a single criterion.

Next slide. If you have three criteria on which you want stakeholders to judge different research topics you will end up having to do 18 pair-wise comparisons as you go through the math. The question is what about more topics and what if you have more criteria? As I hinted to earlier in the presentation we found that there were a number of research topics that stakeholders were interested in and they tended to be far more than four research topics.

Next slide. Here are, for example, the criteria that we found our stakeholders used when identifying research priorities. They focused, in part, on uncertainty about the effectiveness of therapies. There may be lots of evidence about the efficacy of therapies when studied in research studies but less so about the effectiveness of those same therapies when used in the real world. They also examined, they were also interested in the impact of

those therapies on patient-centered outcomes. They worried or they were interested in the quality of the evidence in efficacy studies before deciding whether those therapies are ready for effectiveness studies. They wanted to know about the variability in care across real world settings. That is how much practice variation existed. They worried about cost. They also were interested about the effectiveness, the feasibility, I should say, of conducting effectiveness studies and whether or not such studies would help inform care in diverse setting so would this be a one off situation where you learn about the effectiveness of therapies in a particular setting or was this a broad applicability that could then be translated nationwide, for example. The stakeholders had a range of criteria, in this case, seven different criteria that they thought were important.

Next slide. In this case I am giving you an example of what happens if you have seven criteria and nine topics of interest, which is a representation of what actually happened. You will find that you will need to have stakeholders conduct a series of pair-wise comparisons, far more than three or six or ten.

Next slide. It turns out that in analytic hierarchy process if you are doing these pair-wise comparisons that for each criterion there would be [n(n-1)]/2 pair-wise comparisons where n is the number of research topics being compared and for 9 topics and 7 criteria that turns out to be 252 comparisons. If it takes you about a minute in order to thoughtfully compare two different topics for a particular criterion and then you needed to do this for 9 topics and 7 criteria it will take you about 252 minutes or four solid hours.

Next slide. What we found is that while we engaged stakeholders and they are very excited to participate with us in identifying research priorities they were a bit reluctant to spend four hours straight with no bathroom breaks in order to participate with us. We ended up voted during the conference to actually adopt a modified version of the analytic hierarchy, essentially, to first triage their research topics and then through subsequent discussions then fully deploy AHP. What we ended up doing with our stakeholders is rather than going through each of the seven criteria we ended up asking the stakeholders to

rate the relative importance of the various topics on the overall importance rather than focusing specifically on each of the criteria because, frankly, of feasibility issues.

Next slide. What we found is that topics 1 through 9 using AHP can be rank ordered with respect to the relative importance of those various topics and what you are finding here, for example, is that topic 1 had a normalized priority of 0.22, which essentially means that topic 1, that stakeholders told us that topic 1 had 22% of the importance of all of the 9 topics and if you rank ordered by normalized priority scores you find that topic 9 was felt to be least important, only attracting 2% of the importance of all the topics. You can rank order the topics by normalized priority. What we found in the variability of preferences across stakeholders, just as we had found using the other approaches to rank ordering priorities, is that stakeholders again expressed variable preferences, as you can see here by the relatively wide interquartile ratios. What AHP allows you to do is to quantify and rank order the relative priorities but you are still left with this finding that stakeholders have varying preferences, even with all of this work where you are bringing in stakeholders to discuss the various research topics. I think this is going to be one of the findings that people are going to find engaging stakeholders to rank order priorities is that stakeholders view the research world in different ways. They tend to have different interests with respect to what research questions ought to be addressed.

Next slide. What are CONCERT's reflections on analytic hierarchy process for setting CER priorities? What we find is that the analytic hierarchy process is, indeed, a quite robust method to quantitatively gauge stakeholder preferences for research and when fully deployed can link it very well to the research criteria or to the criteria employed by the stakeholders but that it is not practical when there is a relatively large number of topics or if there is a relatively large number of criteria that stakeholders employ. What I have shown you here is that if you have 9 topics and 7 criteria, which is the example that I reviewed earlier, you will end up requiring stakeholders to conduct a series of 252 pair-wise comparisons, which

depending on time and attention may be impractical to deploy. If you have 5 topics and 5 criteria you can see that you end up with 50 comparisons. If you have even a smaller number of topics and smaller number of criteria then you can clearly reduce the number of pair-wise comparisons required. Our experience is that AHP we ended up using a pragmatic version of the analytic hierarchy process and that we suggest to other groups that you use a similar approach to triage topics before fully deploying analytic hierarchy process because it is a relatively intensive procedure that requires a substantial amount of commitment of time and stakeholder interest. One way to do this would be after the initial triaging that you would then focus on the highest scoring topics and the most important criteria and fully deploying the AHP. Given the variation in preferences that we found using three different ways of eliciting stakeholder priority for research that CONCERT has taken the view that this provides an opportunity for our group to work with different stakeholder groups across different sets of CER priorities. In essence, I doubt very much that we will ever find a single most important topic that all the stakeholders view as being important for study but instead will have a pipeline that is rank ordered and that depending on which stakeholder group you wish to engage in this provides a way to triage those discussions.

Next slide. Next slide.

MS. MOHR: Sorry. We are…

DR. KRISHNAN: Okay. the next slide is I will just say is relatively simple and just wanted to acknowledge the CONCERT consortium received funding from the Agency for Healthcare Research and Quality to conduct this work and that it represents a multi-institutional consortium that is stretching across the U.S. from coast to coast. Thank you very much.

MS. MOHR: Thank you, Jerry. We did receive one question during your talk and I do want to remind people that you can submit questions during people's talks and that will help speed things along. People can submit questions right now. I just in the interest of time would like to make sure that your questions are specific to Jerry's talk and that we will be having more questions at the end of the whole

session. But Jerry the questions that we got from one of the participants is that your list of stakeholders seems to include only those who are professionally involved and not people with COPD and the participant wanted to know how might the process and results change if patients were included?

DR. KRISHNAN: Sure. That's a very good question so we found we use a very deliberate process to engage a wide range of stakeholder groups and I didn't really spend time discussing or reviewing all of the different stakeholders involved. we did have patients with COPD engage in the process. What I presented on the slide was one of the patient advocacy groups. This idea, though, that identifying research priorities may change depending on who is at the table I think is an important one and I think that there hasn't been enough work done in this area about determining what is the right mix of patients and providers and professional societies and researchers and this is an area that I think there is a substantial amount of interest in defining some standards. I think in the United States, for example, there is a new funding agency called the PCORI, Patient-centered Outcomes Research Institute, that has really pushed the boundaries and engaging patients in setting research priorities and developing research studies and this is a fertile area for lots of work at this point.

MS. MOHR: Great. Jerry, actually I am going to ask you one more question, which may take you a little bit of time but just one more and then we'll save all the other questions until the end. We did get a question from a participant can you randomize stakeholders to a subset of pair-wise comparisons to address the issue of responded burden for a large number of pair-wise comparisons?

DR. KRISHNAN: That is a very good question so that approach has been used in other fields where if the total amount of pair-wise comparisons is large and that is in quotes. It depends on your resources and time. That you could randomly allocate different groups to answer different questions. the issues, though, will need to be that you need to have adequate representation of the various stakeholder groups for each of the comparisons that need to be made so you will have to have a very large pool of stakeholders and stakeholder representatives that are participating. The short answer to that question is it is

certainly possible to be done but you will need an even larger number of stakeholder organizations engulfed.

MS. MOHR: Great. Maarten also commented on your presentation, Jerry, about he was thinking maybe you could consider an extra level of sub-criteria in your structure to prioritize endpoints and avoid so many comparisons so maybe at the end we can return to those and talk a little bit more. We have a couple of other questions but, again, I am going to hold those questions until the end. We will move on to a polling question again. This time we want to know using techniques such as AHP to quantify stakeholder preferences would improve the topic prioritization process? That is your opinion. Do you strongly agree with this, disagree, neither agree nor disagree, agree or strongly agree? Go ahead and submit your responses to this and we will see what people think. Again, this is very good news. It looks like the majority of people agree or strongly agree so congratulations, Jerry, on that good presentation. I would like to move on now to Reed Johnson and thank you. Reed?

The Use of Conjoint Analysis to Elicit Patient Preferences in Selecting Treatment Endpoints

F. REED JOHNSON, PhD: Thank Penny and thanks for the community forum for the opportunity to share some of what our group at our Research Triangle Institute has learned about the potential for quantifying patient preferences for health outcomes.

Some of the questions you have asked come up in the context of this kind of research, as well. I have maybe some answers for some of those questions, as well.

Penny, next slide, Penny gave us a quote from the Institute of Medicine about the importance of comparative effectiveness research in helping to make informed decisions. Another feature of the quote was that we are in need to help make informed decisions when there are comparisons between benefits and harms of alternative interventions involved.

Next slide. Comparative effectiveness requires identifying relevant endpoints but relevant for whom? Patients are

important stakeholders but they often have little influence. Their concerns are mediated primarily through physicians, sometimes through patient advocacy groups, but it is uncertain how representative of these stakeholder expressions are so our goal here is to think about how we might measure the concerns of well defined patient stakeholder populations.

Next slide. I would like to deal with a little bit of confusion here about different types of self-reported data. All three of these approaches have played a role in cost effectiveness analysis and maybe have a potential role but we have less experience in comparative effectiveness research. PRO is most familiar. I don't really have to say anything about that. It is just clinical data that can't be directly observed. Don't really get any information about subjective importance to patients. Qualities are used in some health systems but have well known problems and haven't been very popular in the U.S. I am going to focus on this third type of patient-reported or self-reported data but this is data obtained from controlled experiments. I will talk about how we set those up. the instruments, unlike the other two kinds of sources of patient data, require a tailored or application-specific instrument. We don't have a validated instrument that is widely accepted and used. That instrument gives us quantitative instruments of preference weights and from those preference weights we can do a number of things and I will give a few examples in a minute but among those things are healthy time equivalents, maximum acceptable risk, minimum acceptable benefit and willingness to pay or money equivalent value.

Next slide. The idea of preference utility is a concept that is used in every area of theoretical and applied economists except health. Health-state utility is limited to the severity and duration of specific health state conditions but preference utility or what non-health economists just call utility depends on that and everything else that affects human welfare. It provides a very rich conceptual framework for understanding patient concerns and patient behavior and it is really the basis for conjoint or discrete choice experiments.

Next slide. There is some terminological confusion here. Maarten used the term conjoint analysis and it is widely

used and particularly in market research. It is just a conflation of the two words consider jointly but doesn't really explain very much. A term or a label that is gaining wider acceptance and I think is going to win out here, it is commonly used in Europe and it is increasingly winning out in the literature, is discrete choice experiments. I don't like the lab rat implications of that term. My preference is stated choice surveys, which I think explains what we do but that one is not going to win.

Next slide. Here is a choice experiment. This approach was originally developed for market research. It was later refined by psychologists, economists and statisticians. It's based on the idea that we can desegregate objects of choice into compounded parts. This is exactly the same idea as Maarten's and Jerry's presentation. The total value of the choice alternative depends on the values of these component features. In conjoint analysis or discrete choice experiments we simulate a decision context involving hypothetical health endpoints. Then we analyze the resulting pattern of choice responses statistically to get estimates of the implicit decision weights that people were using that were consistent with the choices that we observed so this is an indirect approach rather than using the methods that the other two speakers talked about that require respondents to actually give you those numbers. In conjoint analysis nobody gives you any numbers. they just give you choices. In order to make this work it requires a wide variety of general survey research skills, knowledge of experimental design techniques and some expertise in advanced statistical analysis.

Next slide, please. This is an example of a benefit/risk tradeoff question or a preference tradeoff task. The idea is to offer two alternatives, which are described by a number of different features or characteristics or attributes. We have two efficacy attributes, pain and stiffness, and this is for osteoarthritis. We have a mild to moderate side effect, GI problems and two serious side effect risks, bleeding ulcer and risk of heart attack or stroke. The idea here is to explain to patients to provide them with sufficient background information to make them sufficiently informed for them to be able to process and give us a meaningful choice here. the challenge of knowing how informed is informed is a major problem in this kind of

research. We describe the features carefully in simple definitions in sixth grade English. We get into big arguments with clinicians about accuracy of technical clinical concepts sort of translated into sixth grade English. We are concerned about over conditioning our subjects. That is we can't learn much about the behavior or wild seals from observing trained seals so we are not interested in making these, our patients, clinical experts. We want to make them sufficiently informed. Maybe as informed as they might be if they made a reasonable amount of effort in looking online, reading brochures, talking to their physician. We give them a simple risk tutorial to help them understand basic concepts. We show risk as a count percentage and a risk with graphing. There are serious risk communication challenges here. our goal is to control the psychological stimuli but engage subjects in realistic decision problems so that they can actually answer the question what would you choose if you were actually faced with these alternatives?

Next slide. Take a look at an example here on adherence. The columns here are preference weights for two levels of glucose control rescaled between zero and one. This utility is an ordinal scale in this methodology. There is no true zero so we can rescale this vertical axis. Everything I show you will be scaled differently and it doesn't matter because all we are interested in is relative changes so in this case an improvement from satisfactory to best glucose control is about three-fourths as important as best control. That is the way to kind of read these results. Specifically, improving glucose control from satisfactory to best improves preference utility by 0.28.

Next slide. Now, let's compare glucose control with number of injections. Remember, they didn't give us these numbers. We derived these numbers from a series of choice questions. Now, the problems is that if it takes two injections a day to achieve best control we are going to have a problem. the decrease in preference utility of moving from one injection to two injections is 0.61. That is about twice as much as the increase in utility when moving from satisfactory to best control so if it is using near normal glucose control requires an additional injection patients with preferences like these will be non-adherent because non-adherence makes them, adherence makes

them subjectively worse off. Now, this offers some opportunities for thinking about how we might improve outcomes. It might require better diabetes education if we believe patients don't understand the long-term risks of suboptimal control or a better strategy might b finding technologies that reduce the discomfort of injection.

Next slide. Here we have a study on physician versus patient preferences for hepatitis B. I am going to skip over this since Maarten gave us a similar example. We find similar opportunities here for dissidence between physicians and patients. We find dissidence between two countries, Germany and Turkey and within each country we find dissidence between what the relative importance is.

Next slide. Next slide. For German patients and German physicians the least important attribute was weight of evidence. For Turkish patients and physicians they also agreed on the least important but everybody disagrees about most important, both between patients and physicians and between countries.

Next slide. Following up on the question of the role of these kinds of approaches for comparing benefits versus risk so let's suppose we have, again, charts that show the relative importance of various outcomes and now let's put a level of individual endpoints.

Next slide. Next slide. The gain for going from five months to ten months in progression free survival in renal cell carcinoma is 0.84. Now, let's take that arrow that goes from five months to ten months, turn it upside down starting at zero risk.

Next slide. How far would we have to increase the risk of liver failure to take away that 0.84 gain in benefit from going from five months to ten months in progression free survival and the answer is that patients would tolerate up to a little more than 2% increase in the chance of liver failure as part of the therapeutic risk. We can do exactly the same calculation that is to calibrate one outcome in terms of another, which basically solves the problem of comparing dissimilar outcomes or any continuous attribute. We can get equivalent weighting time, we can get equivalent time in ill health, which would be a generalized quality, we can get equivalent money value. We did a study last

year that did an equivalent gain in weight for an improvement in efficacy. This is the solution or a solution to the non-comparability problem of benefit/risk, cost effectiveness and comparative effectiveness analysis.

Next slide. Now, here is another experiment that we did a few years ago on vasomotor symptoms. Vasomotor symptoms are related to hormone replacement therapy. All subjects were given both absolute risk and relative risk levels in the background information. Half were shown absolute and half were shown relative risk as a tradeoff question. This chart compares the three levels of symptom relief. In each case, risk tolerance measures as a maximum acceptable risk was greater for the absolute risk treatment than for the relative risk treatment. This is more or less a standard result.

Next slide. We can compare this to the actual risk from the original Women's Health Initiative analysis. This has been redone many times since. It turns out in this case that the qualitative results are not much affected by whether risks are shown as relative or absolute. It turns out that it did matter for MI risks, however.

Next slide. We have a number of methodological challenges. Critics of this method will say, well, you ask people a hypothetical question you will get a hypothetical answer. We try to overcome that as best we can by making questions as realistic as possible and that helps to mitigate hypothetical bias but it is true that a number of things can influence how valid or reliable our result are depending on the actual experience people have had with the condition, the temptation to give a socially acceptable response and there are ways of sort of calibrating or testing the degree of hypothetical bias using actual behavior versus a state of behavior. Jerry talked about the cognitive challenges associated with how much effort is required. We try to reduce those cognitive challenges in our surveys by using effective low-level descriptions of critical endpoints. There are all kinds of problems with how to help people understand surrogate markers. People are generally innumerate so those concepts are difficult and then there are some methodological challenges for which there is limited consensus among researchers. I don't have time to share all the dirty laundry here.

Next slide. I guess my view is that if we are going to effectively incorporate patient perspective in the protocol development we need to quantify things. It is not good enough just to come up with relative rankings or ask patient advocacy groups to weigh in on a decision in a qualitative way. We really need to treat patient preferences as a source of evidence. This is an unusual but novel idea for most clinicians that we can actually quantify patient preferences and use that as evidence as we do clinical evidence but these methods do offer methods for quantifying the relative values of endpoints. They have good validity and reliability for relatively simple tradeoff problems but we have lots of interesting research challenges in applying these approaches to more difficult problems.

MS. MOHR: Thank you very much, Reed. That was excellent. We have a very tough question for you, actually, that is specific to your talk and then we have a wide variety of general questions that have come in and I am inviting people to submit their questions while people are answering right now but the specific question for you, Reed, is how do you manage to get patient input on things like treatment for renal cell carcinoma or breast cancer without unacceptable distress at the questioning process? I am thinking about the - - review - - comparisons to someone currently undergoing the distress of treatment.

DR. JOHNSON: Okay. That is an excellent question and it is one that we were just dealing with last week for melanoma, I guess. There are a couple of things. One is, of course, that these people are very sick and nevertheless their preferences should count here and we also have some concerns about whether healthcare is really responsive to the concerns of endstage patients. We have administered these questionnaires quite successfully while people are lying on an infusion table. They are eager to talk about what they are worried about. I am not talking about, of course, severely ill patients but patients who are quite ill. The case of cancer in general we often use caregiver surrogates. These caregivers often are people who are making or helping to make therapeutic decisions. They are close to the patient. We have gotten pretty good success in comparing caregiver versus patient preferences. We get

much more alignment there than we do between patients and physicians of patients.

MS. MOHR: Okay. Thank you very much. There is a question here that actually I think is related to more Maarten's presentation about what are the main differences between AHP and DCE and I am not sure that was something that Maarten went through and I know we had a lot of people joining us a little bit later but I am wondering the question is do both methods measure the preferences and if yes can both methods be used in parallel? Maarten, I am going to put that back to you since that is a major important one for the audience today.

DR. JOHNSON: We will see if Maarten says the same thing I would say.

MS. MOHR: Okay.

DR. IJZERMAN: Thank you. I think this is also a very tough question, which is not easy to answer in a couple of minutes. Both methods, both approaches they come from different schools of thought, operator research or marketing and they also have different theoretical backgrounds and underpinnings so there is quite some debate about the question whether they measure preferences in a utility framework. I think most people would agree that if you take a decision support approach, if that will be the main objective to support a decision of a team or a group of people or stakeholders then a technique like AHP really would be sufficient and helpful because it makes the decision very specific and transparent. On the other hand, if you want to collect preferences over a large number of people and a wide number of alternatives then most people would say conjoint analysis measures preferences in - - theory. The big question is do because in AHP and in MCDA you also use a kind of a value approach. You have to value different options and in conjoint analysis you use choice modeling. You offer a scenario, a real life, a realistic scenario, whereas, in MCDA you have decomposed scenarios and pairs comparisons so the question is if the AHP is more realistic in that perspective than conjoint analysis. The debate here is that if you look at the empirical comparisons between AHP and MCDA techniques and conjoint analysis and there is nothing clear difference in terms of the predictive validity of consumer choice so it is not

very easy to say which method falls better in the prediction of the actual choice that people make, which is actual the real preference, which Reed was talking about, so that would be my first part of the answer. Maybe Reed and Jerry can comment on the last one.

DR. KRISHNAN: No. That's a good answer. Really good.

DR. JOHNSON: I would add a couple of things but in the interest of time I think that's a really good answer.

MS. MOHR: Okay. I also wanted to ask this is a specific question for Jerry that we weren't able to get to earlier but Jerry what is lost by exploring priorities that do not rise to the top? Might they be important to special populations?

DR. KRISHNAN: Thank you so much for the question. I think this is a key issue here that just because a particular research topic did not rise to the top doesn't mean that it does not merit study. For example, particularly for patients that have rare conditions it might impact a relatively small number of people and they had a relatively modest financial footprint compared to conditions that affect a larger number of people but it doesn't necessarily mean that those kinds of topics don't merit study. That is the reason why, at least as we employ AHP for chronic obstructive pulmonary disease, we didn't say that some topics are not important to us for study but rather this provides a rank order to have some way to decide which sets of topics would be studies you would want to engage in at the start versus those you take in later. The other approach, though, is as I mentioned earlier different stakeholder groups had different sets of preferences for what ought to be topics that they would feel are most important and having AHP allows you to understand how those stakeholder preferences were influenced by different criteria and provides a method by which our researchers could engage different sets of stakeholders depending on the research question at hand and so the, I guess, my short answer to this, my summary of my response to this question is that I don't think we believe that topics that are lower on the list are unimportant but this is one way to rank order the topics and it doesn't prohibit groups from studying topics that are not at the top of the list.

MS. MOHR: Okay. Thank you. I had a—

DR. JOHNSON: [Interposing] Can I very quickly—

MS. MOHR: Go ahead, Reed.

DR. JOHNSON: Can I very quickly followup on a question that was asked earlier on sample size, how many question you need and all that sort of things?

MS. MOHR: Yes.

DR. JOHNSON: This is a dimensionality problem in stated preference research that in order to answer the question you have to ask people too many questions or you need very large numbers of people. This is an experimental design problem. there is a report coming out on best practice in experiment design tat will appear as a task force report that will appear in Value in Health. Last year there was a similar task force report on best practices or good practice guidance in general for conjoint analysis. I would refer anyone interested to start maybe with those two reports to answer the question about these practical question about how do you actually collect the data.

MS. MOHR: Thank you. We are actually starting to get flooded with questions right now, which is fantastic so we are not going to have time to answer all of them or pose all of them. What I would suggest is that potentially we can send these by email to our presenters and we will send back to all the participants the answers to these questions so that we can make sure we get them answered. But just a couple of ones that I want to get to before the end of our webcast today. This is for everybody and I think I will start out with Maarten. Are there any specific software available for either AHP or conjoint analysis?

DR. IJZERMAN: Okay, well, thank you for that question. It is quite easy to answer, although, it is quite difficult to list all the software packages within the timeframe of this panel. There is quite a few software packages for MCDA approaches and AHP in particular. I can send an overview, a list of software packages and web links if people are interested. I would refer to Decision Lens, which is actually the software that is being developed by the original developer, Thomas Saaty and his family. Decision Lens is one of the examples but there are many more

packages that also produce AHP. I even found an app on the app store producing AHP scores so there is quite a bit of software on that.

In conjoint analysis there is also quite a bit of software. We tend to work with Sawtooth Software but even the bigger statistical packages provide you the excremental design options like SBSS they will give you external design figures for designing conjoint experiments. For full software packages we can provide a list. That's no problem.

MS. MOHR: Great. That's fantastic. Also, we did get a question about whether or not we could get a list of references and I know we have a few that we put together for this webinar but if maybe the presenters can send us some of their favorites we can share that with participants. There is just time for a couple more questions and one of them is how do you weight cost considerations for healthcare providers and insurers while still considering efficacy and patient preferences? I guess I will just open that up to the panel whoever speaks first.

DR. JOHNSON: I mean this is the argument about whether we should be using cost effectiveness or cost benefit analysis. In conjoint analysis all you have to do is include cost as one of the attributes and you can get the same kinds of tradeoffs I showed for maximum acceptable risk or essentially maximum acceptable cost or what some people call willingness to pay. That deals with the threshold problem in cost effectiveness analysis where we get an actual data to determine what the threshold should be.

MS. MOHR: Reed, I have another question for you and that is have you compared the three choice questions among patients during treatment and the same patients looking back and is there any dissidence?

DR. JOHNSON: Well, we actually have just completed a survey on MS that deals with that. Actually, we have done it and we have gotten inconsistent results. It does seem to be that for chronic progressive conditions patients are pretty good at predicting, their preferences are pretty stable with

respect to differences in severity. For more acute conditions the results are much less reliable.

MS. MOHR: Okay. Thank you very much. I think we are nearing the end of our webcast today and with that I'd like to thank everyone for your participating and thank you very much to the panel. It has been an excellent conversation and thank you very much for all the participants and the great questions you have sent us. We are still even getting questions as I speak and so as I mentioned before we will be sending out some answers to the questions we weren't able to get to today and I thank you and wish you all a good day.

[END RECORDING]